

# Unorganized Malicious Attacks Detection

**Ming Pang**

PANGM@LAMDA.NJU.EDU.CN

*National Key Laboratory for Novel Software Technology  
Nanjing University  
Nanjing 210023, China*

**Wei Gao**

GAOW@LAMDA.NJU.EDU.CN

*National Key Laboratory for Novel Software Technology  
Nanjing University  
Nanjing 210023, China*

**Min Tao**

TAOM@NJU.EDU.CN

*Department of Mathematics  
Nanjing University  
Nanjing 210023, China*

**Zhi-Hua Zhou**

ZHOUSH@LAMDA.NJU.EDU.CN

*National Key Laboratory for Novel Software Technology  
Nanjing University  
Nanjing 210023, China*

**Editor:**

## Abstract

Recommender system has attracted much attention during the past decade, and many attack detection algorithms have been developed for better recommendation. Most previous approaches focus on the shilling attacks, where the attack organizer fakes a large number of user profiles by the same strategy to promote or demote an item. In this paper, we study a different attack style: *unorganized malicious attacks*, where attackers respectively use a small number of user profiles to attack their own target items without any organizer. This attack style occurs in many real applications, yet relevant study remains open. In this paper, we formulate the unorganized malicious attacks detection as a variant of matrix completion problem, and prove that attackers can be detected theoretically. We propose the Unorganized Malicious Attacks detection (UMA) algorithm, which can be viewed as a proximal alternating splitting augmented Lagrangian method. We verify, both theoretically and empirically, the effectiveness of our proposed algorithm.

**Keywords:** attacks detection, recommender systems, alternating direction method, augmented Lagrangian method

## 1. Introduction

Online activities have been an essential part in our daily life as the flourish of the Internet, e.g., increasing customers prefer shopping on Amazon and eBay; Lots of people enjoy watching different movies and TV shows on Youtube and Netflix, etc. There is a big challenge to recommend suitable products effectively as the number of users and items increases drastically; therefore, various collaborative filtering techniques have been developed in diverse

systems so as to help customers choose their favorite products in a set of items (Bresler et al., 2014; Li et al., 2009; Rao et al., 2015).

Many collaborative filtering approaches are vulnerable to spammers and manipulations of ratings (Gunes et al., 2014; Ling et al., 2013), and attackers may bias systems by inserting fake rating scores into the user-item rating matrix. Some attackers may increase the popularity of their own items (push attack) while some others may decrease the popularity of their competitors' items (nuke attack). Most attack detection studies focus on shilling attacks and show good detection performance on kinds of shilling attack strategies (Hurley et al., 2009; Ling et al., 2013; Mehta, 2007). They consider the situation that all the fake user profiles are produced by the same strategy to promote or demote a particular item. For example, an attack organizer may produce hundreds of fake users by a strategy that each fake users gives high scores to the most popular movies and low scores to the target movie to demote it.

Various techniques have been developed to control shilling attacks, e.g., online sites require real names and telephone numbers for registrations; CAPTCHA is used to determine that the response is not generated by a robot; customers are allowed to rate a product after purchasing this product on the shopping website. Based on these measures, traditional shilling attacks may suffer high cost. For example, small online sellers in e-commerce like Amazon might not be willing to fake hundreds of customer rating profiles to implement a shilling attack.

In this paper, we investigate a new attacks model named *unorganized malicious attack*, where attackers respectively use a small number of user profiles to attack their own targets without any organizer. This attack style happens in many real applications, e.g., online sellers on Amazon may fake a few customer rating profiles to nuke their competitors' high-quality shoes; writers may hire several readers to give high scores to their low-quality books. In fact, it has been shown that systems are seriously affected by small amounts of unorganized malicious attacks.<sup>12</sup>

We first formulate the unorganized malicious attacks detection as a variant of matrix completion problem. Let  $X$  denote the ground-truth rating matrix without attacks and noises, and the matrix is low-rank since the users' preferences are affected by several factors (Salakhutdinov et al., 2007). Let  $Y$  be the sparse attack-score matrix, and  $Z$  denotes a noisy matrix. What we can observe is a (or partial) matrix  $M$  such that  $M = X + Y + Z$ . As far as we know, previous works do not make similar formulation for attack detection. The main difference between our optimization problem and robust PCA (Candès et al., 2011) is that robust PCA focuses on recovering low-rank part  $X$  from complete or incomplete matrix and we pay more attention to distinguishing the sparse attack term  $Y$  from the small perturbation noise term  $Z$ .

Theoretically, we prove that the low-rank rating matrix  $X$  and the sparse matrix  $Y$  can be recovered under some classical matrix-completion assumptions, but with a different optimization problem. We propose the Unorganized Malicious Attacks detection (UMA) algorithm, which can be viewed as a proximal alternating splitting augmented Lagrangian method. Some new techniques have been developed to prove its global convergence. Finally,

---

1. <http://www.forbes.com/sites/suwcharmananderson/>.

2. <http://how-to-post-fake-reviews-on-amazon.blogspot.com/>.

experimental results verify the effectiveness of our proposed algorithm in comparison with the state-of-the-art methods of attack detection.

The rest of this paper is organized as follows. Section 2 reviews some related works. Section 3 introduces the framework of unorganized malicious attacks, and Section 4 proposes our UMA algorithm. Section 5 presents theoretical justification for attack detection, matrix recovery and convergence of UMA algorithm. Section 6 shows our experiments, and Section 7 concludes this work.

## 2. Related Work

Collaborative filtering (CF) is one of the most successful techniques to build recommender systems. The core assumption of CF is that if users express similar interests in the past, they will share common interest in the future (Goldberg et al., 1992). Significant progress about CF has been made since then (Bresler et al., 2014; Li et al., 2009; Rao et al., 2015; Salakhutdinov et al., 2007). There are two main categories of conventional CF (based on the user-item rating matrix) which are memory-based and model-based CF algorithms. Memory-based CF predicts a user’s rating on an item based on the entire or part of the user-item matrix. It can be subdivided into user-based and item-based CF. A typical user-based CF approach predicts the ratings of a user by aggregating the ratings of some similar users. User similarity is defined by a similarity metric, usually the cosine similarity or the Pearson correlation (Singhal, 2001). Many modifications and adjustments about the similarity metric have been proposed (Adomavicius and Singhal, 2005; Zhang and Pu, 2007). Item-based CF approaches predict the rating of an item for a user according to the ratings of items the user has given (Deshpande and Karypis, 2004).

Model-based CF approaches use the user-item matrix to train prediction models and recommendations are generated from the prediction models (Ekstrand et al., 2011). For example, the mixture model learns the probability distribution of items in each clusters (Kleinberg and Sandler, 2008); Matrix factorization techniques learn latent factors of users and items from the user-item matrix and then use the low-rank approximation matrix to predict the score of unrated items; From probabilistic perspective, Salakhutdinov and Mnih (2008a) propose probabilistic matrix factorization framework. Considering about side information besides the user-item matrix, many works expand the CF paradigm (Basilico and Hofmann, 2004; Salakhutdinov et al., 2007; Li et al., 2009).

However the two main categories of CF schemes are both vulnerable to attacks (Gunes et al., 2014; Ling et al., 2013). Increasing attention has been given to attack detection. Researchers have proposed several kinds of methods which can be mainly thought as statistical, clustering, classification and data reduction-based methods (Gunes et al., 2014). These methods mainly focus on shilling attacks where the attack organizer fakes a large number of user profiles by the same strategy to promote or demote a particular item. Statistical methods are used to detect anomalies who give suspicious ratings. Hurley et al. (2009) propose a Neyman-Pearson statistical attack detection method to distinguish attackers from normal users. Similarly, probabilistic Bayesian network models are used in Li and Luo (2011). Based on attributes derived from user profiles, classification methods (Mobasher et al., 2009) detect attacks by kNN, SVM, rough set theory, etc.

An unsupervised clustering algorithm based on several classification attributes (Bryan et al., 2008) is presented in Bhaumik et al. (2011). They apply  $k$ -means clustering based on these attributes and classify users in the smallest cluster as attackers. Instead of using traditional nearest neighbor methods, Mehta (2007) proposes a PLSA-based clustering method. Mehta and Nejd1 (2009) propose the variable selection method, which treats users as variables and calculates their covariance matrix. By analyzing the principal components of the covariance matrix, those users with the smallest coefficient in the first  $l$  principal components are chosen in the final variable selection. Ling et al. (2013) try to predict the users' ratings by using a low-rank matrix factorization method and classify users with the lowest reputation as attackers.

These methods implement detection based on the attack strategy about how the fake user rating profiles are produced. When recommender systems are under unorganized malicious attacks, different attackers adopt different strategies to produce fake user rating profiles or hire existing users to attack. The traditional attack detection methods may be not suitable in this case.

### 3. The Formulation

In this section, we introduce the general form of an attack profile, and then we give a detailed comparison of unorganized malicious attacks compared to shilling attacks. The formal definition of unorganized malicious attacks and the corresponding detection problem formulation are also presented.

#### 3.1 Notations

We begin with some notations used throughout this paper. Let  $\|X\|$ ,  $\|X\|_F$  and  $\|X\|_*$  denote the operator norm, Frobenius norm and nuclear norm of matrix  $X$ , respectively. Let  $\|X\|_1$  and  $\|X\|_\infty$  be the  $\ell_1$  and  $\ell_\infty$  norm of matrix  $X$  seen as vectors, respectively. Further, we define the Euclidean inner product between two matrices as  $\langle X, Y \rangle := \text{trace}(XY^\top)$ , where  $Y^\top$  means the transposition of  $Y$ . Therefore, We have  $\|X\|_F^2 = \langle X, X \rangle$ .

Let  $P_\Omega$  denote an operator of linear transformation which acts on matrices space, and we also denote  $P_\Omega$  by the linear space of matrices supported on  $\Omega$  when it is clear from the context. Then,  $P_{\Omega^\top}$  represents the space of matrices supported on  $\Omega^c$ . For an integer  $m$ , let  $[m] := \{1, 2, \dots, m\}$ .

#### 3.2 Problem Formulation

The general form of an attack profile is shown in Figure 1 which is first defined by Bhaumik et al. (2006). the selected items,  $I_S$ , are rated by the rating function  $\delta$ . The filler items of  $I_F$  are selected randomly and rated by the function  $\theta$ . The function  $\Upsilon$  determines the rating of the target item  $i_t$ . The remaining items are unrated.

In each kind of shilling attacks, the target item  $i_t$  is fixed; the number of rated items ( $k + l$ ) is fixed; the rating functions are fixed. For example, in random attacks, the filler items  $I_F$  are randomly chosen and the function  $\theta$  gives  $I_F$  the mean score of the system. In bandwagon attacks, the selected items  $I_S$  are chosen from the popular items and the function gives  $I_S$  a high score. The patterns of attack profiles are the same in each kind of

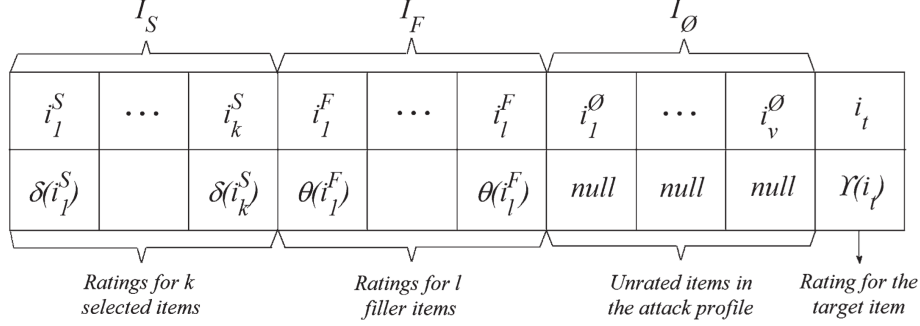


Figure 1: General form of an attack profile.

shilling attack. Besides, the number of attack profiles requires to be large in the common settings of shilling attacks.

However, in unorganized malicious attacks, the number of rated items, the target item and the rating functions are not constrained to be the same. Besides, we assume that each attackers produces a small number of attack profiles. The formal definition of unorganized malicious attacks is given in the following.

Let  $U_{[m]} = \{U_1, U_2, \dots, U_m\}$  and  $I_{[n]} = \{I_1, I_2, \dots, I_n\}$  denote  $m$  users and  $n$  items, respectively. Let  $X \in \mathbb{R}^{m \times n}$  be the rating matrix.  $X_{ij}$  denotes the score that user  $U_i$  gives item  $I_j$  without any attack or noise, i.e.,  $X_{ij}$  reflects the ground-truth feeling of user  $U_i$  on item  $I_j$ . Suppose that the highest score is  $R_{max}$ , then  $0 < X \leq R_{max}$ . Throughout this work, we assume that  $X$  is a low-rank matrix as in classical matrix completion (Salakhutdinov and Mnih, 2008b). The intuition behind this assumption is that only a few factors influence users' preferences.

Usually, the ground-truth matrix  $X$  may be corrupted by a system noisy matrix  $Z$ . For example, if  $X_{ij} = 4.8$  for  $i \in [m]$ , then, it is acceptable that user  $U_i$  gives item  $I_j$  score 5 or 4.6. In this paper, we consider the independent Gaussian noise, i.e.,  $Z = (Z_{ij})_{m \times n}$  where each element  $Z_{ij}$  is drawn i.i.d. from the Gaussian distribution  $\mathcal{N}(0, \sigma)$  with parameter  $\sigma$ .

Let  $M$  be the observed matrix which is corrupted by random noise and unorganized malicious attacks. We define an unorganized malicious attack with respect to a user set  $U_K$  ( $K \subset [m]$ ) if  $|K| \leq \iota$  and  $|M_{ij} - X_{ij}| \geq \epsilon$  for some  $j \in [n]$  and every  $i \in K$ . The parameter  $\kappa$  controls the number of users and parameter  $\epsilon$  is used to distinguish malicious users from the normal. Intuitively, unorganized malicious attacks mean that attackers respectively use a small number of user profiles to attack their own targets.

It is necessary to distinguish unorganized malicious attacks from noise. Generally speaking, user  $U_i$  gives item  $I_j$  a normal score if  $|M_{ij} - X_{ij}|$  is very small, while user  $U_i$  makes an attack to item  $I_j$  if  $|M_{ij} - X_{ij}| > \epsilon$ . For example, if the ground-truth score of item  $I_j$  is 4.9 for user  $U_i$ , then user  $U_i$  makes a noisy rating by giving  $I_j$  score 5, yet makes an attack by giving  $I_j$  score 2. Therefore, we assume that  $\|Z\|_F \leq \delta$ , where  $\delta$  is a small parameter.

Let  $Y = M - X - Z = (Y_{ij})_{m \times n}$  be the malicious-attack-score matrix. Then,  $Y_{ij} = 0$  if user  $U_i$  does not attack item  $I_j$ , otherwise  $|Y_{ij}| \geq \epsilon$ . We assume that  $Y$  is a sparse matrix, and the intuition behind this assumption is that each malicious user attacks a small number

of items. Notice that we can not directly recover  $X$  and  $Y$  from  $M$  because such recovery is an NP-Hard problem (Candès et al., 2011). We consider the following optimization problem.

$$\begin{aligned} \min_{X,Y,Z} \quad & \|X\|_* + \tau\|Y\|_1 + \alpha\langle X, Y \rangle \\ \text{s.t.} \quad & X + Y + Z = M, \\ & \|Z\|_F \leq \delta, \end{aligned} \tag{1}$$

where the term  $\langle X, Y \rangle$  is introduced to control the impact of  $Y$  in the nuke attack detection. The intuition behind this term is that the malicious rating bias  $Y_{ij}$  and the true rating  $X_{ij}$  are always opposite in nuke attack, i.e.  $X_{ij}Y_{ij} \leq 0$ . So we need to minimize  $\langle X, Y \rangle$  to better distinguish sparse matrix  $Y$  and noisy matrix  $Z$ . The third term is  $\alpha\langle (X - R_{max}), Y \rangle$  in the push attack detection.

In many real applications, we can not get a full matrix  $M$ , and only partial entries can be observed. Let  $\Omega \in [m] \times [n]$  be the set of observed entries. We define an orthogonal projection  $P_\Omega$  onto the linear space of matrices supported on  $\Omega \subset [m] \times [n]$  as follows.

$$P_\Omega M = \begin{cases} M_{ij} & \text{for } (i, j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

Our final optimization framework for unorganized malicious nuke attack detection can be formulated as follows.

$$\begin{aligned} \min_{X,Y,Z} \quad & \|X\|_* + \tau\|Y\|_1 + \alpha\langle X, Y \rangle \\ \text{s.t.} \quad & P_\Omega(X + Y + Z) = P_\Omega M, \\ & \|P_\Omega(Z)\|_F \leq \delta. \end{aligned} \tag{2}$$

There have been many works (Candès et al., 2011; Feng et al., 2013; Mackey et al., 2011; Peng et al., 2012) on recovering low-rank part  $X$  from complete or incomplete matrix. However, we pay more attention to distinguishing the sparse noise term  $Y$  from the small perturbation term  $Z$ . In order to find nonzero entries of  $Y$ , a new term  $\langle X, Y \rangle$  is added which leads to a more challenging optimization task and gets a much better performance. Further details about the proposed approach, theoretical analysis and experiments are explained below in Section 4, 5, 6.

#### 4. The Proposed Approach

It is rather difficult to optimize the formulations of Eqns. (1) and (2) with theoretical guarantees, because this optimization includes three-block non-convex programming with coupled objective function  $\alpha\langle X, Y \rangle$ . The difficulties lie in three hands: i) the objective function is non-convex due to the existence of coupled term; ii) the subgradients of the involved three functions are non-Lipschitz; and iii) there are three blocks variables involved. Therefore, we consider the following perturbation formulation inspired by Cai et al. (2010);

Candès et al. (2011).

$$\begin{aligned} \min_{X,Y,Z} \quad & \|X\|_* + \tau\|Y\|_1 + \alpha\langle X, Y \rangle + \frac{\kappa}{2}\|X\|_F^2 + \frac{\kappa}{2}\|Y\|_F^2 \\ \text{s.t.} \quad & X + Y + Z = P_\Omega M, \\ & Z \in \mathbf{B}, \\ & \mathbf{B} := \{Z \mid \|P_\Omega(Z)\|_F \leq \delta\}, \end{aligned}$$

where  $\kappa > 0$  is a regularization parameter. As  $\kappa \rightarrow 0$ , this formulation degenerates into Eqn. (2). We further get the augmented Lagrangian function as follows.

$$\begin{aligned} \mathcal{L}(X, Y, Z, \Lambda, \beta) := & \|X\|_* + \tau\|Y\|_1 + \alpha\langle X, Y \rangle \\ & + \frac{\kappa}{2}\|X\|_F^2 + \frac{\kappa}{2}\|Y\|_F^2 - \langle \Lambda, L \rangle + \frac{\beta}{2}\|L\|_F^2, \end{aligned}$$

where  $L = X + Y + Z - P_\Omega M$  and  $\beta$  is a positive constant.

It is noteworthy that there is a coupling term in our objective function, and some traditional algorithms (He et al., 2015; Tao and Yuan, 2011) can not be applied directly. We propose a proximal alternating splitting augmented Lagrangian method to solve this optimization, which inherits the advantages of ASALM (Tao and Yuan, 2011). We will provide global convergence guarantee for this algorithm in Section 5.

More specifically, let  $\gamma > 0$  and  $(X^k, Y^k, \Lambda^k)$  is given. We first update

$$Z^{k+1} = \arg \min_{Z \in \mathbf{B}} \mathcal{L}_A(X^k, Y^k, Z, \Lambda^k, \beta),$$

and it is easy to get the closed form solution

$$Z_{ij}^{k+1} = \begin{cases} \min\{1, \delta/\|P_\Omega N\|_F\} N_{ij} & \text{if } (i, j) \in \Omega \\ N_{ij} & \text{otherwise.} \end{cases} \quad (3)$$

where  $N = \frac{1}{\beta}\Lambda^k + P_\Omega M - X^k - Y^k$ . Then, we update

$$X^{k+1} = \arg \min_{X \in \mathbb{R}^{m \times n}} \mathcal{L}_A(X, Y^k, Z^{k+1}, \Lambda^k, \beta) + \gamma\beta\|X - X^k\|_F^2/2.$$

Lemma 2 gives the closed solution  $X^{k+1}$  as

$$\mathcal{D}_\mu(\beta\mu(P_\Omega M + \frac{1}{\beta}\Lambda^k - Y^k - Z^{k+1} - \frac{\alpha}{\beta}Y^k + \gamma X^k)), \quad (4)$$

where  $\mu = 1/(\kappa + \gamma\beta + \beta)$  and  $\mathcal{D}_\mu$  is defined by Lemma 2. Further, we update

$$Y^{k+1} = \arg \min_{Y \in \mathbb{R}^{m \times n}} \mathcal{L}_A(X^{k+1}, Y, Z^{k+1}, \Lambda^k, \beta).$$

Lemma 1 gives the closed form solution  $Y^{k+1}$  as

$$\mathcal{S}_{\tau v}(v\beta(P_\Omega M + \frac{1}{\beta}\Lambda^k - Z^{k+1} - X^{k+1}) - v\alpha X^{k+1}), \quad (5)$$

where  $v = 1/(\beta + \kappa)$  and  $\mathcal{S}_{\tau v}$  is defined by Lemma 1. Finally, we update

$$\Lambda^{k+1} = \Lambda^k - \beta(X^{k+1} + Y^{k+1} + Z^{k+1} - P_\Omega M). \quad (6)$$

The pseudocode of the proposed UMA method is given in Algorithm 1.

**Algorithm 1** The UMA Algorithm**Input:** matrix  $M$  and parameters  $\tau, \alpha, \beta, \delta, \kappa$  and  $\gamma$ .**Output:** Label vector  $[y_1, \dots, y_m]$  where  $y_i = 1$  if user  $U_i$  is a malicious user; otherwise  $y_i = 0$ .**Initialize:**  $Y^0 = X^0 = \Lambda^0 = 0$ ,  $y_i = 0$  ( $i = 1, \dots, m$ ),  $k = 0$ **Process:**

- 1: **while** not converged **do**
- 2:   Compute  $Z^{k+1}$  by Eqn. (3).
- 3:   Compute  $X^{k+1}$  by Eqn. (4).
- 4:   Compute  $Y^{k+1}$  by Eqn. (5).
- 5:   Update the Lagrange multiplier  $\Lambda^{k+1}$  by  $\Lambda^k - \beta(X^{k+1} + Y^{k+1} + Z^{k+1} - P_\Omega M)$ .
- 6:    $k = k + 1$ .
- 7: **end while**
- 8: if  $\max(|Y_{i,:}|) > 0$ , then  $y_i = 1$  ( $i = 1, \dots, m$ ).

**5. Theoretical Analysis**

In this section, we first prove that attacks can be detected theoretically. Then we give a convergence analysis of our algorithm UMA and prove its global convergence.

**5.1 Detection Guarantees**

We begin with two useful lemmas for the deviation of our proposed algorithms as follows.

**Lemma 1** (*Bruckstein et al., 2009*) For  $\tau > 0$  and  $T \in \mathbb{R}^{m \times n}$ , the closed solution of  $\min_Y \tau \|Y\|_1 + \|Y - T\|_F^2/2$  is given by matrix  $\mathcal{S}_\tau(T)$  with  $(\mathcal{S}_\tau(T))_{ij} = \max\{|T_{ij}| - \tau, 0\} \cdot \text{sgn}(T_{ij})$ , where  $\text{sgn}(\cdot)$  means the sign function.

**Lemma 2** (*Cai et al., 2010*) For  $\mu > 0$  and  $Y \in \mathbb{R}^{m \times n}$  with rank  $r$ , the closed solution of  $\min_X \mu \|X\|_* + \|X - Y\|_F^2/2$  is given by

$$\mathcal{D}_\mu(Y) = S \text{diag}(\mathcal{S}_\mu(\Sigma)) D^\top$$

where  $Y = S \Sigma D^\top$  denotes the singular value decomposition of  $Y$ , and  $\mathcal{S}_\mu(\Sigma)$  is defined in Lemma 1.

For simplicity, our theoretical analysis focuses on square matrix, and it is easy to generalize our results to the general rectangular matrices. Let the singular value decomposition of  $X_0 \in \mathbb{R}^{n \times n}$  be given by

$$X_0 = S \Sigma D^\top = \sum_{i=1}^r \sigma_i s_i d_i^\top$$

where  $r$  is the rank of matrix  $X_0$ ,  $\sigma_1, \dots, \sigma_r$  are the positive singular values, and  $S = [s_1, \dots, s_r]$  and  $D = [d_1, \dots, d_r]$  are the left- and right-singular matrices, respectively. For  $\mu > 0$ , we assume

$$\begin{aligned} \max_i \|S^\top e_i\|^2 &\leq \mu r/n, \\ \max_i \|D^\top e_i\|^2 &\leq \mu r/n, \\ \|SD^\top\|_\infty^2 &\leq \mu r/n^2. \end{aligned} \tag{7}$$



We now present our first main result as follows.

**Theorem 3** *Suppose that the support set of  $Y_0$  be uniformly distributed for all sets of cardinality  $k$ , and  $X_0$  satisfies the incoherence condition given by Eqn. (7). Let  $X$  and  $Y$  be the solution of optimization problem given by Eqn. (1) with parameter  $\tau = O(1/\sqrt{n})$  and  $\alpha = O(1/n)$ . For some constant  $c > 0$  and sufficiently large  $n$ , the following holds with probability at least  $1 - cn^{-10}$  over the choice on the support of  $Y_0$*

$$\|X_0 - X\| \leq \delta \text{ and } \|Y_0 - Y\|_F \leq \delta$$

if  $\text{rank}(X_0) \leq \rho_r n / \mu / \log^2 n$  and  $k \leq \rho_s n^2$ , where  $\rho_r$  and  $\rho_s$  are positive constant.

**Proof** Let  $\Omega$  be the space of matrices with the same support as  $Y_0$ , and let  $T$  denote the linear space of matrices

$$T := \{SX^* + YD^*, X, Y \in \mathbb{R}^{n \times r}\}.$$

We will first prove that, for  $\|P_\Omega P_T\| \leq 1/2$ ,  $(X_0, Y_0)$  is the unique solution if there is a pair  $(W, F)$  satisfying

$$SD^* + W = \tau(\text{sgn}(Y_0) + F + P_\Omega K) \quad (8)$$

where  $P_T W = 0$  and  $\|W\| \leq 1/2$ ,  $P_\Omega F = 0$  and  $\|F\|_\infty \leq 1/2$  and  $\|P_\Omega K\|_F \leq 1/4$ . Notice that  $SD^* + W_0 + \alpha Y_0$  is an arbitrary subgradient of  $\|X\|_* + \alpha \langle X, Y \rangle$  at  $(X_0, Y_0)$ , and  $\tau(\text{sgn}(Y_0) + F_0) + \alpha X_0$  is an arbitrary subgradient of  $\tau\|Y\|_1 + \alpha \langle X, Y \rangle$  at  $(X_0, Y_0)$ . For any matrix  $H$ , we have, by the definition of subgradient,

$$\begin{aligned} & \|X + H\|_* + \tau\|Y - H\|_1 + \alpha \langle X + H, Y - H \rangle \\ & \geq \|X_0\|_* + \tau\|Y_0\|_1 + \alpha \langle X_0, Y_0 \rangle + \alpha \langle Y_0 - X_0, H \rangle \\ & \quad + \langle SD^* + W_0, H \rangle - \tau \langle \text{sgn}(Y_0) + F_0, H \rangle. \end{aligned} \quad (9)$$

By setting  $W_0$  and  $F_0$  satisfying  $\langle W_0, H \rangle = \|P_{T^\perp} H\|_*$  and  $\langle F_0, H \rangle = -\|P_{\Omega^\perp} H\|_1$ , we have

$$\begin{aligned} & \langle SD^* + W_0, H \rangle - \tau \langle \text{sgn}(Y_0) + F_0, H \rangle \\ & = \|P_{T^\perp} H\|_* + \tau\|P_{\Omega^\perp} H\|_1 + \langle SD^* - \tau \text{sgn}(Y_0), H \rangle \\ & = \|P_{T^\perp} H\|_* + \tau\|P_{\Omega^\perp} H\|_1 + \langle \tau(F + P_\Omega K) - W, H \rangle \\ & \geq \frac{1}{2}(\|P_{T^\perp} H\|_* + \tau\|P_{\Omega^\perp} H\|_1) + \tau \langle P_\Omega D, H \rangle \end{aligned} \quad (10)$$

where the second equality holds from Eqn. (8), and the last inequality holds from

$$\langle \tau F - W, H \rangle \geq -|\langle W, H \rangle| - |\langle \tau F, H \rangle| \geq -(\|P_{T^\perp} H\|_* + \tau\|P_{\Omega^\perp} H\|_1)/2$$

for  $\|W\| \leq 1/2$  and  $\|F\|_\infty \leq 1/2$ . We further have

$$\langle \tau P_\Omega K, H \rangle \geq -\frac{\tau}{4}\|P_{\Omega^\perp} H\|_F - \frac{\tau}{2}\|P_{T^\perp} H\|_F \quad (11)$$

from  $\|P_\Omega K\|_F \leq 1/4$  and

$$\begin{aligned} \|P_\Omega H\|_F &\leq \|P_\Omega P_T H\|_F + \|P_\Omega P_{T^\perp} H\|_F \\ &\leq \|P_\Omega P_{T^\perp} H\|_F + \|H\|_F/2 \\ &\leq (\|P_\Omega H\|_F + \|P_{\Omega^\perp} H\|_F)/2 + \|P_\Omega P_{T^\perp} H\|_F. \end{aligned}$$

Combing with Eqns. (9) to (11), we have

$$\begin{aligned} &\|X + H\|_* + \tau\|Y - H\|_1 + \alpha\langle X + H, Y - H \rangle \\ &\geq \|X_0\|_* + \tau\|Y_0\|_1 + \alpha\langle X_0, Y_0 \rangle + \alpha\langle Y_0 - X_0, H \rangle \\ &\quad + \frac{1-\tau}{2}\|P_{T^\perp} H\|_* + \frac{\tau}{4}\|P_{\Omega^\perp} H\|_1 \end{aligned}$$

From the conditions that  $\Omega \cap T = \{0\}$ ,  $\tau = O(1/\sqrt{n})$  and  $\alpha = O(1/n)$ , we have

$$\alpha\langle Y_0 - X_0, H \rangle + \frac{1-\tau}{2}\|P_{T^\perp} H\|_* + \frac{\tau}{4}\|P_{\Omega^\perp} H\|_1 > 0$$

for sufficient large  $n$ . Therefore, we can recover  $X_0$  and  $Y_0$  if there is a pair  $(W, F)$  satisfying Eqn. (8), and the pair  $(W, F)$  can be easily constructed according to Candès et al. (2011). We complete the proof from the condition  $\|Z\|_F \leq \delta$ .  $\blacksquare$

Similarly to the proof of Theorem 3, we present the following theorem for the minimization problem of Eqn. (2).

**Theorem 4** *Suppose that  $X_0$  satisfies the incoherence condition given by Eqn. (7), and  $\Omega$  is uniformly distributed among all sets of size  $m \geq n^2/10$ . We assume that each entry is corrupted independently with probability  $\tau$ . Let  $X$  and  $Y$  be the solution of optimization problem given by Eqn. (2) with parameter  $\tau = O(1/\sqrt{n})$  and  $\alpha = O(1/n)$ . For some constant  $c > 0$  and sufficiently large  $n$ , the following holds with probability at least  $1 - cn^{-10}$*

$$\|X_0 - X\|_F \leq \delta \text{ and } \|Y_0 - Y\|_F \leq \delta$$

if  $\text{rank}(X_0) \leq \rho_r n / \mu / \log^2 n$  and  $\tau \leq \rho_s$ , where  $\rho_r$  and  $\rho_s$  are positive constants.

## 5.2 Convergence Analysis

Before starting to show the convergence, we derive its optimality condition of (3). Let  $\mathcal{W} := \mathbf{B} \times \mathcal{R}^{m \times n} \times \mathcal{R}^{m \times n} \times \mathcal{R}^{m \times n}$ . It follows from Corollaries 28.2.2 and 28.3.1 of Rockafellar (1970) that the solution set of (3) is non-empty. Then, let  $W^* = ((Z^*)^\top, (X^*)^\top, (Y^*)^\top, (\Lambda^*)^\top)^\top$  be a saddle point of (3). It is easy to see that (3) is equivalent to finding  $W^* \in \mathcal{W}$  such that  $\forall W = (Z^\top, X^\top, Y^\top, \Lambda^\top)^\top \in \mathcal{W}$ ,

$$\begin{cases} (Z - Z^*)^\top (-\Lambda^*) \geq 0, \\ \|X\|_* - \|X^*\|_* + (X - X^*)^\top (\kappa X^* + \alpha Y^* - \Lambda^*) \geq 0, \\ \tau\|Y\|_1 - \tau\|Y^*\|_1 + (Y - Y^*)^\top (\alpha X^* + \kappa Y^* - \Lambda^*) \geq 0, \\ X^* + Y^* + Z^* - M = 0, \end{cases} \quad (12)$$

or, in a more compact form  $\text{VI}(\mathcal{W}, \Psi, \theta)$ ,

$$\theta(U) - \theta(U^*) + (W - W^*)^\top \Psi(W^*) \geq 0, \quad \forall W \in \mathcal{W}, \quad (13a)$$

where

$$U = \begin{pmatrix} Z \\ X \\ Y \end{pmatrix}, \quad \theta(U) = \|X\|_* - \|X^*\|_* + \tau\|Y\|_1 - \tau\|Y^*\|_1, \quad (13b)$$

$$\text{and } W = \begin{pmatrix} Z \\ X \\ Y \\ \Lambda \end{pmatrix}, \quad V = \begin{pmatrix} X \\ Y \\ \Lambda \end{pmatrix}, \quad \Psi(W) = \begin{pmatrix} -\Lambda \\ \kappa X + \alpha Y - \Lambda \\ \alpha X + \kappa Y - \Lambda \\ X + Y + Z - M \end{pmatrix}. \quad (13c)$$

Note that  $U$  collects all the primal variables in (12) and it is a sub-vector of  $W$ . Moreover, we use  $W^*$  to denote the solution set of  $\text{VI}(\mathcal{W}, \Psi, \theta)$  and define  $V^* = ((X^*)^\top, (Y^*)^\top, (\Lambda^*)^\top)^\top$  and  $\mathcal{V}^* := \{V^* | W^* \in \mathcal{W}\}$ .

**Remark 5** *Note there is a coupling term in the objective function of the problem (3). The VASALM in Tao and Yuan (2011); He et al. (2015) can not be applied directly. Because, the splitting methods developed in Tao and Yuan (2011); He et al. (2015) with global convergence aims to solve the separable convex programming. Therefore, these existing methods can not be applied to solve problem (3) with global convergence. Hence, we propose a new splitting method, i.e., UMA which aims to solve the problem (3) with coupling objective function. Moreover, UMA inherits the advantage of ASALM in Tao and Yuan (2011), i.e., it updates three blocks variables  $Z$ ,  $X$  and  $Y$  in Gauss-Seidel manner. The proximal term introducing in  $X$ -subproblem is to ensure the global convergence of UMA.*

In the following, we concentrate on the convergence of the proposed UMA. We first prove some properties of the sequence generated by the proposed UMA, which play crucial roles in the coming convergence analysis.

**Lemma 6** *Let  $\{W^k\}$  be generated by UMA. Then, we have the following inequality:*

$$\begin{aligned} \theta(U^{k+1}) - \theta(U^*) + (W^{k+1} - W^*)^\top \Psi(W^*) \\ \geq (\kappa - \alpha)(\|X^{k+1} - X^*\|_F^2 + \|Y^{k+1} - Y^*\|_F^2). \end{aligned} \quad (14)$$

**Proof** First, due to (13c), we have

$$\begin{aligned} & (\Delta W)^\top (\Psi(W^{k+1}) - \Psi(W^*)) \\ &= (\Delta W)^\top \begin{pmatrix} 0 & 0 & 0 & -I \\ 0 & \kappa I_m & \alpha I_m & -I \\ 0 & \alpha I_m & \kappa I_m & -I \\ I & I & I & 0 \end{pmatrix} (\Delta W) \\ &\geq (\kappa - \alpha)(\|X^{k+1} - X^*\|_F^2 + \|Y^{k+1} - Y^*\|_F^2), \end{aligned} \quad (15)$$

where  $\Delta W = W^{k+1} - W^*$ . Consequently,

$$\begin{aligned}
 & \theta(U^{k+1}) - \theta(U^*) + (W^{k+1} - W^*)^\top \Psi(W^{k+1}) \\
 &= \theta(U^{k+1}) - \theta(U^*) + (W^{k+1} - W^*)^\top \Psi(W^*) \\
 &\quad + (W^{k+1} - W^*)^\top (\Psi(W^{k+1}) - \Psi(W^*)) \\
 &\geq (W^{k+1} - W^*)^\top (\Psi(W^{k+1}) - \Psi(W^*)) \\
 &\geq (\kappa - \alpha)(\|X^{k+1} - X^*\|_F^2 + \|Y^{k+1} - Y^*\|_F^2).
 \end{aligned}$$

The first and the second inequalities follow from (13a), (15), respectively. Thus, the conclusion follows directly.  $\blacksquare$

**Lemma 7** *Let  $\{W^k\}$  be generated by UMA. Then, we have the following inequality:*

$$\begin{aligned}
 & \theta(U) - \theta(U^{k+1}) + (W - W^{k+1})^\top \Psi(W^{k+1}) \\
 &+ (W - W^{k+1})^\top \begin{pmatrix} \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ 0 \end{pmatrix} \\
 &\geq \langle V - V^{k+1}, \tilde{G}(V^k - V^{k+1}) \rangle, \quad \forall W \in \mathcal{W}.
 \end{aligned} \tag{16}$$

where the matrix

$$\tilde{G} = \begin{pmatrix} (\gamma + 1)\beta I_m & -\alpha I_m & 0 \\ \beta I_m & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}. \tag{17}$$

**Proof** First, based on the optimal condition (12) and the update scheme of  $\Lambda^{k+1}$ , we have

$$\begin{cases} \langle -\Lambda^{k+1} + \beta[X^k - X^{k+1} + Y^k - Y^{k+1}], Z - Z^{k+1} \rangle \geq 0; \\ \|X\|_* - \|X^{k+1}\|_* + \langle X - X^{k+1}, \alpha Y^k + \kappa X^{k+1} \\ -\Lambda^{k+1} + \beta(Y^k - Y^{k+1}) + \gamma\beta(X^{k+1} - X^k) \rangle \geq 0; \\ \tau\|Y\|_1 + \langle Y - Y^{k+1}, \alpha X^{k+1} + \kappa Y^{k+1} - \Lambda^{k+1} \rangle - \tau\|Y^{k+1}\|_1 \geq 0; \\ \langle X^{k+1} + Y^{k+1} + Z^{k+1} - M - \frac{1}{\beta}(\Lambda^k - \Lambda^{k+1}), \Lambda - \Lambda^{k+1} \rangle \geq 0; \end{cases} \quad \forall W \in \mathcal{W}. \tag{18}$$

Consequently,

$$\begin{cases} \langle -\Lambda^{k+1} + \beta[X^k - X^{k+1} + Y^k - Y^{k+1}], Z - Z^{k+1} \rangle \geq 0; \\ \|X\|_* - \|X^{k+1}\|_* + \langle X - X^{k+1}, \alpha Y^{k+1} + \kappa X^{k+1} - \Lambda^{k+1} \rangle + \langle X - X^{k+1}, \\ (1 + \gamma)\beta(X^{k+1} - X^k) + \beta(X^k - X^{k+1} + Y^k - Y^{k+1}) \rangle \geq 0; \\ \tau\|Y\|_1 + \langle Y - Y^{k+1}, \alpha X^{k+1} + \kappa Y^{k+1} - \Lambda^{k+1} \rangle - \tau\|Y^{k+1}\|_1 \geq 0; \\ \langle X^{k+1} + Y^{k+1} + Z^{k+1} - M - \frac{1}{\beta}(\Lambda^k - \Lambda^{k+1}), \Lambda - \Lambda^{k+1} \rangle \geq 0; \end{cases} \quad \forall W \in \mathcal{W}.$$

Finally, combining the definitions of matrix  $\tilde{G}$  in (17),  $\theta(U)$  in (13b) and  $\Psi(W)$  in (13c), the assertion follows directly.  $\blacksquare$

In the following, we show another important property of UMA.

**Lemma 8** *Let  $\{W^k\}$  be generated by UMA. Then, it holds that*

$$\langle Y^k - Y^{k+1}, \alpha(X^k - X^{k+1}) \rangle + \kappa \|Y^k - Y^{k+1}\|_F^2 \leq \langle \Lambda^k - \Lambda^{k+1}, Y^k - Y^{k+1} \rangle. \quad (19)$$

**Proof** By setting  $Y = Y^k$  in the third inequality of (18) we get

$$\tau \|Y^k\|_1 - \tau \|Y^{k+1}\|_1 + \langle Y^k - Y^{k+1}, \alpha X^{k+1} + \kappa Y^{k+1} - \Lambda^{k+1} \rangle \geq 0.$$

Similarly, taking  $k := k - 1$  and  $Y = Y^{k+1}$  in the third inequality of (18) we have

$$\tau \|Y^{k+1}\|_1 - \tau \|Y^k\|_1 + \langle Y^{k+1} - Y^k, \alpha X^k + \kappa Y^k - \Lambda^k \rangle \geq 0.$$

By adding the above two inequalities, we complete the proof. ■

**Theorem 9** *Let  $\{W^k\}$  be generated by UMA. Assume that  $\alpha < 1$  in model (3) and  $\beta > 0$ ,  $\gamma > 0$  in Algorithm (1). Then, we have the following contractive property:*

$$\|V^{k+1} - V^*\|_G^2 \leq \|V^k - V^*\|_G^2 - \Delta_{k+1}, \quad (20)$$

where

$$\Delta_{k+1} := \frac{\epsilon_1}{2\beta} \|\Lambda^k - \Lambda^{k+1}\|_F^2 + \epsilon_2 \|Y^k - Y^{k+1}\|_F^2 + \zeta \|X^k - X^{k+1}\|_F^2, \quad (21)$$

$$\zeta = \left( \frac{\gamma + 1}{2} \beta - \frac{\beta}{2(1 - \epsilon_1)} - \frac{\alpha}{4(\frac{\beta}{2} + \kappa - \frac{\alpha}{4(\kappa - \alpha)} - \epsilon_2)} - \frac{\beta^2}{4(\kappa - \alpha)} \right),$$

and the matrix

$$G = \begin{pmatrix} (\gamma + 1)\beta I_m & 0 & 0 \\ 0 & \beta I_m & 0 \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix},$$

and the positive scalars  $\epsilon_{1,2}$  are sufficient small.

**Proof** First, note that

$$\begin{aligned} & \left\langle W^* - W^{k+1}, \begin{pmatrix} \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ 0 \end{pmatrix} \right\rangle \\ &= \langle M - X^{k+1} - Y^{k+1} - Z^{k+1}, \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \rangle \\ &= \langle \Lambda^{k+1} - \Lambda^k, X^k - X^{k+1} + Y^k - Y^{k+1} \rangle. \end{aligned} \quad (22)$$

The first and second equalities follows from  $X^* + Y^* + Z^* - M = 0$  and the  $\Lambda^{k+1}$ -updating scheme of (6). Then, by setting  $W = W^*$  in inequality (16) we get

$$\begin{aligned}
 0 &\geq \theta(U^{k+1}) - \theta(U^*) + (W^{k+1} - W^*)^\top \Psi(W^{k+1}) \\
 &\quad + \left\langle W^{k+1} - W^*, \begin{pmatrix} \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ \beta[X^k - X^{k+1} + Y^k - Y^{k+1}] \\ 0 \end{pmatrix} \right\rangle \\
 &\quad + \langle V^* - V^{k+1}, \tilde{G}(V^k - V^{k+1}) \rangle \\
 &\geq (\kappa - \alpha) \|X^{k+1} - X^*\|_F^2 + (\kappa - \alpha) \|Y^{k+1} - Y^*\|_F^2 \\
 &\quad + \langle V^* - V^{k+1}, \tilde{G}(V^k - V^{k+1}) \rangle \\
 &\quad + \langle \Lambda^k - \Lambda^{k+1}, X^k - X^{k+1} + Y^k - Y^{k+1} \rangle.
 \end{aligned} \tag{23}$$

The last inequality follows from (14) and (22).

Next, we analyze the right hand side of (23). Due to (19) and the definitions of matrices  $\tilde{G}$  (17) and  $G$  (9), we have

$$\begin{aligned}
 &(\kappa - \alpha) \|X^{k+1} - X^*\|_F^2 + (\kappa - \alpha) \|Y^{k+1} - Y^*\|_F^2 + \langle V^* - V^{k+1}, \tilde{G}(V^k - V^{k+1}) \rangle \\
 &\quad + \langle \Lambda^k - \Lambda^{k+1}, X^k - X^{k+1} + Y^k - Y^{k+1} \rangle \\
 &\geq (\kappa - \alpha) \|X^{k+1} - X^*\|_F^2 + (\kappa - \alpha) \|Y^{k+1} - Y^*\|_F^2 \\
 &\quad + \langle \Lambda^k - \Lambda^{k+1}, X^k - X^{k+1} \rangle + \langle Y^k - Y^{k+1}, \alpha(X^k - X^{k+1}) \rangle \\
 &\quad + \kappa \|Y^k - Y^{k+1}\|_F^2 + \langle V^* - V^{k+1}, G(V^k - V^{k+1}) \rangle \\
 &\quad + \beta \langle Y^* - Y^{k+1}, X^k - X^{k+1} \rangle - \alpha \langle X^* - X^{k+1}, Y^k - Y^{k+1} \rangle.
 \end{aligned} \tag{24}$$

On the other hand, we have the following identity:

$$\begin{aligned}
 &\langle V^* - V^{k+1}, G(V^k - V^{k+1}) \rangle \\
 &= \frac{1}{2} \|V^{k+1} - V^*\|_G^2 - \frac{1}{2} \|V^k - V^*\|_G^2 + \frac{1}{2} \|V^{k+1} - V^k\|_G^2.
 \end{aligned} \tag{25}$$

From the definition (9), the matrix  $G$  is positive definite matrix.

Then, using Cauchy-Schwarz inequalities, we have

$$\beta \langle Y^* - Y^{k+1}, X^k - X^{k+1} \rangle \geq (\alpha - \kappa) \|Y^{k+1} - Y^*\|_F^2 - \frac{\beta^2}{4(\kappa - \alpha)} \|X^k - X^{k+1}\|_F^2,$$

$$\begin{aligned}
 \langle Y^k - Y^{k+1}, \alpha(X^k - X^{k+1}) \rangle &\geq -\left(\frac{\beta}{2} + \kappa - \frac{\alpha}{4(\kappa - \alpha)} - \epsilon_2\right) \|Y^k - Y^{k+1}\|_F^2 \\
 &\quad - \frac{\alpha}{4\left(\frac{\beta}{2} - \frac{\alpha}{4(\kappa - \alpha)} + \kappa - \epsilon_2\right)} \|X^k - X^{k+1}\|_F^2,
 \end{aligned}$$

$$\langle \Lambda^k - \Lambda^{k+1}, X^k - X^{k+1} \rangle \geq -\frac{1 - \epsilon_1}{2\beta} \|\Lambda^k - \Lambda^{k+1}\|_F^2 - \frac{\beta}{2(1 - \epsilon_1)} \|X^k - X^{k+1}\|^2,$$

$$-\alpha \langle X^* - X^{k+1}, Y^k - Y^{k+1} \rangle \geq -(\kappa - \alpha) \|X^{k+1} - X^*\|^2 - \frac{\alpha}{4(\kappa - \alpha)} \|Y^k - Y^{k+1}\|_F^2.$$

where  $\epsilon_{1,2} > 0$ . Note the last inequality uses  $\alpha < 1$ .

Then, substituting the above four inequalities and (25) into the right-hand-side of (24), we get

$$\begin{aligned} & (\kappa - \alpha)\|X^{k+1} - X^*\|_F^2 + (\kappa - \alpha)\|Y^{k+1} - Y^*\|_F^2 + \langle V^* - V^{k+1}, \tilde{G}(V^k - V^{k+1}) \rangle \\ & + \langle \Lambda^k - \Lambda^{k+1}, X^k - X^{k+1} + Y^k - Y^{k+1} \rangle \\ & \geq \Delta_{k+1} + \frac{1}{2}\|V^{k+1} - V^*\|_G^2 - \frac{1}{2}\|V^k - V^*\|_G^2. \end{aligned}$$

Finally, combining the above inequality and (23), the inequality (20) follows directly.  $\blacksquare$

Based on the above theorem, we have the following corollary immediately.

**Theorem 10** *Let the sequence  $\{V^k\}$  be generated by the proposed UMA. Assume that*

$$\kappa > \alpha \text{ and } \kappa^2 - \alpha\kappa - \frac{\alpha}{4} < 0. \quad (26)$$

$$\beta \in \left( 0, (\kappa - \alpha)\gamma + (\kappa - \alpha)\sqrt{\gamma^2 - \frac{4\alpha}{\kappa^2 - \alpha\kappa - \frac{\alpha}{4}}} \right). \quad (27)$$

Then, we have

1. *The sequence  $\{V^k\}$  is bounded.*
2.  $\lim_{k \rightarrow \infty} \{\|Y^k - Y^{k+1}\|_F^2 + \|X^k - X^{k+1}\|_F^2 + \|\Lambda^k - \Lambda^{k+1}\|_F^2\} = 0.$

**Proof** The first assertion follows from (20) directly. We now prove the second assertion. Since  $\kappa > \alpha$  and  $\beta$  is satisfied with (27), and  $\epsilon_{1,2} \rightarrow 0+$ , it holds that

$$\frac{\gamma + 1}{2}\beta - \frac{\beta}{2(1 - \epsilon_1)} - \frac{\alpha}{4(\frac{\beta}{2} + \kappa - \frac{\alpha}{4(\kappa - \alpha)} - \epsilon_2)} - \frac{\beta^2}{4(\kappa - \alpha)} > 0.$$

Consequently,  $\Delta_{k+1} \geq 0$ , where  $\Delta_{k+1}$  is defined in (21). Then, it follows from (20) that

$$\sum_{k=0}^{\infty} \Delta_{k+1} \leq \|V^0 - V^*\|_G^2 < +\infty,$$

which immediately implies that

$$\begin{aligned} \lim_{k \rightarrow \infty} \|Y^k - Y^{k+1}\|_F &= 0, \\ \lim_{k \rightarrow \infty} \|X^k - X^{k+1}\|_F &= 0, \\ \lim_{k \rightarrow \infty} \|\Lambda^k - \Lambda^{k+1}\|_F &= 0, \end{aligned} \quad (28)$$

i.e., the second assertion.  $\blacksquare$

We are now ready to prove the convergence of the proposed UMA.

**Theorem 11** *Let  $\{V^k\}$  and  $\{W^k\}$  be the sequences generated by the proposed UMA. Assume that the model's parameters  $\alpha, \kappa$  are satisfies with (26) and the penalty parameter  $\beta$  is satisfied with (27). Then, we have*

1. *Any cluster point of  $\{W^k\}$  is a solution point of (12).*
2. *The sequence  $\{V^k\}$  converges to some  $V^\infty \in \mathcal{V}^*$ .*
3. *The sequence  $\{U^k\}$  converges to a solution point of (3).*

**Proof** Because of the assertion (28), it follows from (16) that

$$\theta(U) - \theta(U^{k+1}) + (W - W^{k+1})^\top \Psi(W^{k+1}) \geq 0, \quad \forall W = (Z^\top, X^\top, Y^\top, \Lambda^\top)^\top \in \mathcal{W}. \quad (29)$$

We also have  $\theta(U) - \theta(U^k) + (W - W^k)^\top \Psi(W^k) \geq 0, \quad \forall W = (Z^\top, X^\top, Y^\top, \Lambda^\top)^\top \in \mathcal{W}$ .

Since  $\{W^k\}$  is bounded, it has at least one cluster point. Let  $W^\infty$  be a cluster point of  $\{W^k\}$  and the subsequence  $\{W^{k_j}\}$  converges to  $W^\infty$ . It follows from (29) that

$$\lim_{j \rightarrow \infty} \theta(U) - \theta(U^{k_j}) + (W - W^{k_j})^\top \Psi(W^{k_j}) \geq 0, \quad \forall W = (Z^\top, X^\top, Y^\top, \Lambda^\top)^\top \in \mathcal{W}.$$

This means that  $W^\infty$  is a solution of  $\text{VI}(\mathcal{W}, \Psi, \theta)$ . Then the inequality (20) is also valid if  $V^*$  is replaced by  $V^\infty$ . Therefore, the non-increasing sequence  $\|V^k - V^\infty\|_G^2$  converges to 0 since  $V^\infty$  is a cluster point of  $\{V^k\}$ .

Also, the updating scheme of  $\Lambda^{k+1}$  in (6) implies that

$$Z^{k+1} = M - X^{k+1} - Y^{k+1} + \frac{1}{\beta}(\Lambda^k - \Lambda^{k+1}).$$

Combining the above equality, (28) and  $\lim_{k \rightarrow \infty} \|V^k - V^\infty\|_G^2 = 0$ , we have  $W^k$  converges to  $W^\infty$ . It implies that the sequence  $U^k$  converges to a solution point of (3). Thus, the third assertion holds.  $\blacksquare$

## 6. Experiments

In this section, we compare our proposed approach UMA with the state-of-the art approaches for attack detection.

To evaluate the performance of these methods, we use attacks detection precision, recall and F1 as the performance measures. These performance measures can be calculated as follows (Gunes et al., 2014),

$$\begin{aligned} \text{Precision} &= \frac{\# \text{true positives}}{\# \text{true positives} + \# \text{false positives}} \\ \text{Recall} &= \frac{\# \text{true positives}}{\# \text{true positives} + \# \text{false negatives}} \\ \text{F1} &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \end{aligned}$$

in which  $\#$  true positives is the number of attack profiles correctly detected,  $\#$  false positives is the number of misclassified normal profiles, and  $\#$  false negatives is the number of attack profiles that are missed.



### 6.1 Datasets

We first conduct our experiments on the common-used datasets **MovieLens 100K** and **MovieLens 1M** collected and released by GroupLens.<sup>3</sup> The rating scores are integers from 1 to 5, where 1 and 5 are the worst and best, respectively. Dataset **MovieLens 100K** contains 100000 ratings of 943 users over 1682 movies, and Dataset **MovieLens 1M** contains 1000209 ratings of 6040 users over 3706 movies. It is noteworthy that those datasets do not contain attackers and we have to add simulation attack profiles, which will be discussed in Section 6.3.

We also collect a real dataset **Douban 10K** with 35 attack profiles identified by the Douban website, where registered users record rating information over various films, books, clothes, etc.<sup>4</sup> We gather 12095 ratings of 213 users over 155 items, where rating scores are integers from 1 to 5. Among all the 213 user profiles, 35 profiles are attack profiles.

### 6.2 Comparison Methods and Implementation Details

We compare with the state-of-the-art approaches for attack detection and robust PCA as follows.

- **N-P:** A statistical algorithm which identifies attack profiles based on the Neyman-Pearson statistics (Hurley et al., 2009).
- **k-means:** A cluster algorithm based on classification attributes (Bhaumik et al., 2011).
- **PCAVarSel:** A PCA-based variable selection algorithm by computing covariance between users to locate malicious users (Mehta and Nejd, 2009).
- **MF-based:** A matrix factorization algorithm by computing the reputation scores among users (Ling et al., 2013).
- **RPCA:** A low-rank matrix recovery method considering sparse and small perturbation noise (Candès et al., 2011).

In our experiments, we set  $\tau = 10/\sqrt{m}$  and  $\alpha = 10/m$ . A rating can be viewed as malicious rating if the deviation is greater than 3 from the ground-truth rating since the scale of rating in our systems is between 1 and 5; therefore, we set parameter  $\beta = \tau/3$  as the entries of  $Y$  will be nullified if they are smaller than  $\tau/\beta$  according to Eqn. (5). We finally select  $\delta = \sqrt{mn/200}$ . We use the power method (Halko et al., 2011) to approximate Eqn. (4) so that our method has the same scalability as the power method.

### 6.3 Comparison Results

As mentioned before, the datasets **MovieLens 100K** and **MovieLens 1M** do not contain attackers; therefore, we first use a combination of several traditional attack strategies to construct the datasets with unorganized malicious attacks. These traditional attack strategies include average attack strategy, random attack strategy and bandwagon attack strategy (Lam and

3. <http://grouplens.org/datasets/movielens/>.

4. <http://www.douban.com/>.

	Movielens 100K			Movielens 1M		
	Precision	Recall	F1	Precision	Recall	F1
UMA	<b>0.934±0.003</b>	<b>0.883±0.019</b>	<b>0.908±0.011</b>	<b>0.739±0.009</b>	<b>0.785±0.023</b>	<b>0.761±0.016</b>
RPCA	0.908±0.010	0.422±0.048	0.575±0.047	0.342±0.003	0.558±0.028	0.424±0.009
N-P	0.774±0.015	0.641±0.046	0.701±0.032	0.711±0.007	0.478±0.018	0.572±0.014
k-means	0.723±0.171	0.224±0.067	0.341±0.092	0.000±0.000	0.000±0.000	0.000±0.000
PCAVarSel	0.774±0.009	0.587±0.024	0.668±0.019	0.278±0.007	0.622±0.022	0.384±0.011
MF-based	0.911±0.009	0.814±0.008	0.860±0.009	0.407±0.005	0.365±0.004	0.385±0.005

Table 1: Detection precision, recall and F1 compared with other algorithms on MovieLens 100K and MovieLens 1M which are under unorganized malicious attacks based on a combination of traditional strategies.

	Movielens 100K			Movielens 1M		
	Precision	Recall	F1	Precision	Recall	F1
UMA	<b>0.929±0.013</b>	<b>0.865±0.032</b>	<b>0.896±0.022</b>	<b>0.857±0.005</b>	<b>0.733±0.003</b>	<b>0.790±0.002</b>
RPCA	0.797±0.046	0.659±0.097	0.721±0.097	0.635±0.012	0.391±0.022	0.484±0.015
N-P	0.244±0.124	0.145±0.089	0.172±0.084	0.273±0.020	0.099±0.031	0.144±0.035
k-means	0.767±0.029	0.234±0.042	0.357±0.051	0.396±0.026	0.300±0.039	0.341±0.035
PCAVarSel	0.481±0.027	0.168±0.017	0.248±0.023	0.120±0.006	0.225±0.012	0.157±0.008
MF-based	0.556±0.023	0.496±0.021	0.524±0.022	0.294±0.012	0.264±0.010	0.278±0.011

Table 2: Detection precision, recall and F1 compared with other algorithms on MovieLens 100K and MovieLens 1M which are under general unorganized malicious attacks.

Riedl, 2004). Here, each attacker randomly chooses one strategy to fake the user rating profiles and promote one item whose average rating is lower than 2. In line with the setting of previous attack detection works, we set the filler ratio (percentage of rated items in total items) as 0.01 and the filler items are drawn from the top 10% most popular items. We set the spam ratio (number of attack profiles/number of all user profiles) as 0.2. Table 6.3 shows the experimental results on datasets MovieLens 100K and MovieLens 1M which are under unorganized malicious attacks based on a combination of traditional strategies.

Besides inserting new user profiles into the rating matrix, attackers can hire existing users to attack. We reconstruct datasets from MovieLens 100K and MovieLens 1M as follows. We set the filler ratio as 0.01, and the filler items are drawn from the top 10% most popular items. Most attack profiles (75%) are generated by i) randomly selecting an item with average rating lower than 2, ii) randomly selecting a user profile with a rating lower than 2 to the chosen item, and iii) the selected user profile can be viewed as an attack profile by modifying the rating to 5. The other attack profiles (25%) are produced similar to that of the above paragraph. Table 6.3 demonstrates the comparison results on the simulated datasets from MovieLens 100K and MovieLens 1M.

Table 6.3 shows the experiments on dataset Douban 10K. The experimental results in Table 6.3, 6.3, 6.3 show that our proposed UMA approach achieves the best performance on three measures: Precision, Recall and F1, and our approach takes superiority in all

Methods	UMA	RPCA	N-P	k-means	PCAVarSel	MF-based
Precision	<b>0.800</b>	0.535	0.250	0.321	0.240	0.767
Recall	<b>0.914</b>	0.472	0.200	0.514	0.343	0.657
F1	<b>0.853</b>	0.502	0.222	0.396	0.282	0.708

Table 3: Detection precision, recall and F1 compared with other algorithms on dataset Douban 10K.

datasets. It is easy to observe that traditional attack detection approaches fail to work for unorganized malicious attacks, because those methods depend on the properties of shilling attacks, e.g., the k-means method and N-P method work if the attack profiles are similar in the view of the supervised features or their latent categories, and the PCAVarSel method achieves good performance only if attack profiles have more common unrated items than normal profiles. However, those good properties do not exist in the unorganized malicious attacks.

The RPCA and MF-based methods try to find the ground-truth matrix from the observed matrix, whereas they hardly separate the sparse attack matrix from the noisy matrix, and tend to suffer from low precision, especially on large-scale and heavily sparse datasets.

Since different systems may contain different spam ratios (number of attack profiles/number of all user profiles), we compare our UMA algorithm with other methods by varying the spam ratio from 2% to 20% in Figure 2. Our UMA approach always gets robust and better performance in different spam ratios, whereas the comparison methods (except the RPCA method) achieve worse performance for small spam ratio, e.g., the N-P approach detects almost nothing. Although the RPCA method is as stable as our method UMA in different spam ratios, there is a performance gap between RPCA and UMA which becomes bigger when the dataset gets larger and sparser from MovieLens 100K to MovieLens 1M.

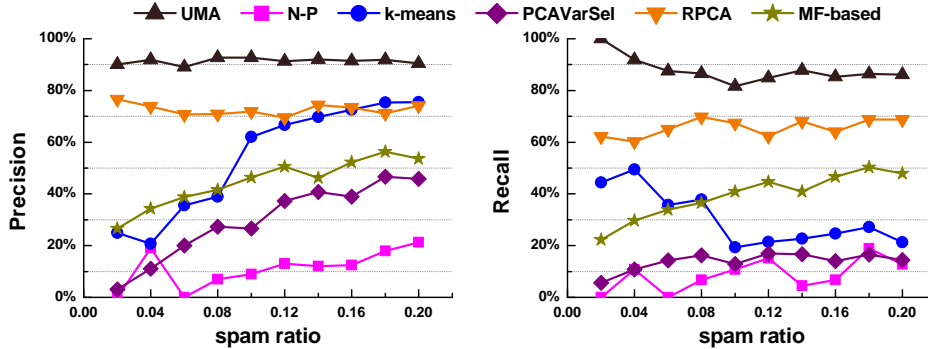


Figure 2: Detection precision and recall on dataset MovieLens 100K under unorganized malicious attacks. The spam ratio (number of attack profiles/number of all user profiles) varies from 0.02 to 0.2.

## 7. Conclusion

The attack detection plays an important role to improve the quality of recommendation in systems, whereas most previous methods focus on shilling attacks. In this paper, we first formulate the unorganized malicious attacks detection as a variant of matrix completion problem. Then we propose the Unorganized Malicious Attacks detection (UMA) algorithm, which can be viewed as a proximal alternating splitting augmented Lagrangian method. We give the proof of global convergence theoretically, and experimental results show that our proposed algorithm achieves significantly better performance than the state-of-the-art approaches for attack detection.

## References

- G. Adomavicius and A. Singhal. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
- J. Basilico and T. Hofmann. Unifying collaborative and content-based filtering. In *Proceedings of the 21st International Conference on Machine Learning*, pages 65–72, 2004.
- R. Bhaumik, C. Williams, B. Mobasher, and R. Burke. Securing collaborative filtering against malicious attacks through anomaly detection. In *Proceedings of the 4th Workshop on Intelligent Techniques for Web Personalization*, 2006.
- R. Bhaumik, B. Mobasher, and R. D. Burke. A clustering approach to unsupervised attack detection in collaborative recommender systems. In *Proceedings of the 7th International Conference on Data Mining*, pages 181–187, 2011.
- G. Bresler, G. Chen, and D. Shah. A latent source model for online collaborative filtering. In *Proceedings of the 28th Advances in Neural Information Processing Systems*, pages 3347–3355, 2014.
- A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.
- K. Bryan, M. O’Mahony, and P. Cunningham. Unsupervised retrieval of attack profiles in collaborative recommender systems. In *Proceedings of the 2nd ACM Conference on Recommender Systems*, pages 155–162, 2008.
- J.-F. Cai, E. J. Candès, and Z.-W. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.
- E. J. Candès, X. D. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011.
- M. Deshpande and G. Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems*, 22(1):143–177, 2004.
- M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, 4(2):81–173, 2011.

- J.-S. Feng, H. Xu, and S.-C. Yan. Online robust pca via stochastic optimization. In *Proceedings of the 27th Advances in Neural Information Processing Systems*, pages 404–412, 2013.
- D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
- I. Gunes, C. Kaleli, A. Bilge, and H. Polat. Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review*, 42(4):767–799, 2014.
- N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011.
- B.-S. He, M. Tao, and X.-M. Yuan. A splitting method for separable convex programming. *IMA Journal of Numerical Analysis*, 35(1):394–426, 2015.
- N. J. Hurley, Z. P. Cheng, and M. Zhang. Statistical attack detection. In *Proceedings of the 3rd ACM Conference on Recommender Systems*, pages 149–156, 2009.
- J. Kleinberg and M. Sandler. Using mixture models for collaborative filtering. *Journal of Computer and System Sciences*, 74(1):49–69, 2008.
- S. K. Lam and J. Riedl. Shilling recommender systems for fun and profit. In *Proceedings of the 13th International Conference on World Wide Web*, pages 393–402, 2004.
- B. Li, Q. Yang, and X.-Y. Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th International Conference on Machine Learning*, pages 617–624, 2009.
- C. Li and Z.-G. Luo. Detection of shilling attacks in collaborative filtering recommender systems. In *Proceedings of the 2nd International Conference of Soft Computing and Pattern Recognition*, pages 190–193, 2011.
- G. Ling, I. King, and M. R. Lyu. A unified framework for reputation estimation in online rating systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 2670–2676, 2013.
- L. W. Mackey, M. I. Jordan, and A. Talwalkar. Divide-and-conquer matrix factorization. In *Proceedings of the 25th Advances in Neural Information Processing Systems*, pages 1134–1142, 2011.
- B. Mehta. Unsupervised shilling detection for collaborative filtering. In *Proceedings of the 22nd International Conference on Artificial Intelligence*, pages 1402–1407, 2007.
- B. Mehta and W. Nejdl. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User-Adapted Interaction*, 19(1-2):65–97, 2009.
- B. Mobasher, R. Burke, R. Bhaumik, and J. J. Sandvig. Attacks and remedies in collaborative recommendation. *Intelligent Systems*, 22(3):56–63, 2009.

- Y.-G. Peng, A. Ganesh, J. Wright, W.-L. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *Pattern Analysis and Machine Intelligence*, 34(11):2233–2246, 2012.
- N. Rao, H.-F. Yu, P. Ravikumar, and I. Dhillon. Collaborative filtering with graph information: Consistency and scalable methods. In *Proceedings of the 29th Advances in Neural Information Processing Systems*, pages 2098–2106, 2015.
- R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th International Conference on Machine Learning*, pages 880–887, 2008a.
- R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *Proceedings of the 22nd Advances in Neural Information Processing Systems*, pages 1257–1264, 2008b.
- R. Salakhutdinov, A. Mnih, and G. Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning*, pages 791–798, 2007.
- A. Singhal. Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24(4):35–43, 2001.
- M. Tao and X.-M. Yuan. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1):57–81, 2011.
- J.-Y. Zhang and P. Pu. A recursive prediction algorithm for collaborative filtering recommender systems. In *Proceedings of the 1st ACM Conference on Recommender Systems*, pages 57–64, 2007.